ABSTRACT

A method of identifying features for a classifier includes identifying a set of elements that share a common characteristic, and then identifying a subset of elements within that set which share a second characteristic. Features are then selected that are more commonly possessed by the elements in the subset than the elements in the set but excluding the subset, and that are more commonly possessed by the elements in the set but excluding the subset, as compared to the elements outside the set. A further method of identifying features for a classifier includes defining a list of features, selecting a first feature from that list, identifying a set of elements that possess that first feature, and then identifying a subset of elements within that set which possess any other feature. A feature is then selected that is more commonly possessed by the elements in the subset than the elements in the set but excluding the subset, and that is more commonly possessed by the elements in the set but excluding the subset, as compared to the elements outside the set. If this feature is not already in the list of features, it is added to it. Another feature that has not already been selected is chosen from the list, and the process is repeated using this feature. This continues until every feature in the list of features has been selected.